

# yang\_seonhyeHW20

```
library(alr4)

## Loading required package: car
## Loading required package: carData
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

library(data.table)
ais <- data.frame(ais)
```

## Question 1

```
attach(ais, warn.conflicts = F)

fit_intercept <- lm(Bfat~1)
fit <- ~Ht+Wt+LBM+BMI+SSF
add1(fit_intercept, fit, test = "F")

## Single term additions
##
## Model:
## Bfat ~ 1
##          Df Sum of Sq    RSS     AIC   F value   Pr(>F)
## <none>            7701.1 737.45
## Ht      1     272.3 7428.9 732.18    7.3295  0.007370 **
## Wt      1      0.0 7701.1 739.45    0.0000  0.998176
## LBM     1    1008.3 6692.8 711.10   30.1326 1.214e-07 ***
## BMI     1    270.9 7430.2 732.22    7.2921  0.007519 **
## SSF     1    7142.0  559.1 209.64 2554.8760 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit_intercept <- update(fit_intercept, . ~ . + SSF)

add1(fit_intercept, fit, test="F")

## Single term additions
##
## Model:
## Bfat ~ SSF
##          Df Sum of Sq    RSS     AIC F value   Pr(>F)
## <none>            559.09 209.64
## Ht      1     110.36 448.73 167.23  48.940 3.929e-11 ***
## Wt      1     174.40 384.69 136.12  90.216 < 2.2e-16 ***
## LBM     1     210.66 348.43 116.12 120.318 < 2.2e-16 ***
## BMI     1     127.14 431.95 159.53  58.572 8.278e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

fit_intercept <- update(fit_intercept, . ~ . + LBM)
add1(fit_intercept, fit, test="F")

## Single term additions
##
## Model:
## Bfat ~ SSF + LBM
##      Df Sum of Sq   RSS   AIC F value Pr(>F)
## <none>       348.43 116.12
## Ht      1     4.175 344.25 115.69  2.4012 0.1228
## Wt      1    215.814 132.61 -77.01 322.2273 <2e-16 ***
## BMI     1     3.140 345.29 116.29  1.8003 0.1812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit_intercept <- update(fit_intercept, . ~ . + Wt)

add1(fit_intercept, fit, test="F")

## Single term additions
##
## Model:
## Bfat ~ SSF + LBM + Wt
##      Df Sum of Sq   RSS   AIC F value   Pr(>F)
## <none>       132.61 -77.010
## Ht      1     4.7280 127.88 -82.343 7.2833 0.007564 **
## BMI     1     4.0237 128.59 -81.234 6.1644 0.013870 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit_intercept <- update(fit_intercept, . ~ . + Ht)

add1(fit_intercept, fit, test="F")

## Single term additions
##
## Model:
## Bfat ~ SSF + LBM + Wt + Ht
##      Df Sum of Sq   RSS   AIC F value Pr(>F)
## <none>       127.88 -82.343
## BMI     1    0.16417 127.72 -80.603 0.2519 0.6163
#The final model includes SSF, LBM, Wt and Ht

```

## Question 2

```

fitmod <- lm(Bfat~Ht + Wt + LBM + BMI + SSF, data=ais)
drop1(fitmod, test="F")

## Single term deletions
##
## Model:
## Bfat ~ Ht + Wt + LBM + BMI + SSF

```

```

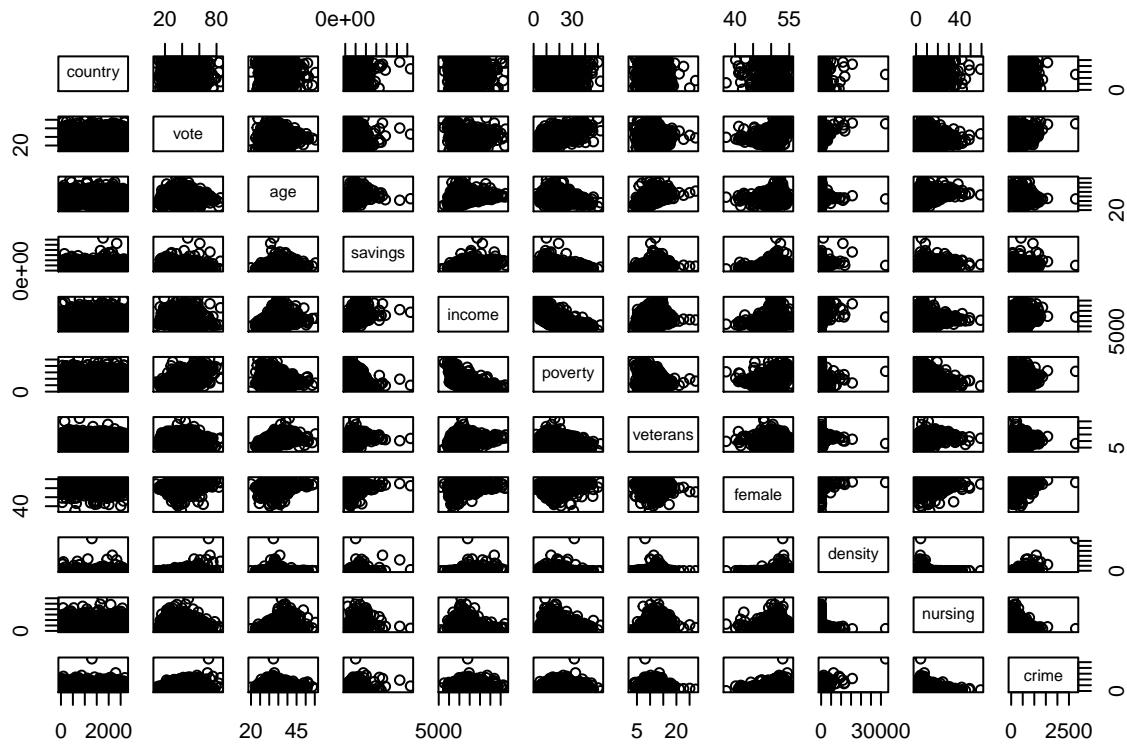
##          Df Sum of Sq    RSS      AIC  F value    Pr(>F)
## <none>            127.72 -80.603
## Ht      1     0.868 128.59 -81.234   1.3327    0.2497
## Wt      1    100.371 228.09  34.538 154.0297 < 2.2e-16 ***
## LBM     1    256.755 384.47 140.009 394.0189 < 2.2e-16 ***
## BMI     1     0.164 127.88 -82.343   0.2519    0.6163
## SSF     1    15.800 143.52 -59.042  24.2472 1.795e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- lm(Bfat~Ht + Wt + LBM + BMI + SSF, data=ais)
drop1(fitmod, test="F")

## Single term deletions
##
## Model:
## Bfat ~ Ht + Wt + LBM + BMI + SSF
##          Df Sum of Sq    RSS      AIC  F value    Pr(>F)
## <none>            127.72 -80.603
## Ht      1     0.868 128.59 -81.234   1.3327    0.2497
## Wt      1    100.371 228.09  34.538 154.0297 < 2.2e-16 ***
## LBM     1    256.755 384.47 140.009 394.0189 < 2.2e-16 ***
## BMI     1     0.164 127.88 -82.343   0.2519    0.6163
## SSF     1    15.800 143.52 -59.042  24.2472 1.795e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . - BMI)
drop1(fitmod, test="F")

## Single term deletions
##
## Model:
## Bfat ~ Ht + Wt + LBM + SSF
##          Df Sum of Sq    RSS      AIC  F value    Pr(>F)
## <none>            127.88 -82.343
## Ht      1     4.728 132.61 -77.010   7.2833  0.007564 **
## Wt      1    216.367 344.25 115.687 333.3050 < 2.2e-16 ***
## LBM     1    256.775 384.66 138.106 395.5524 < 2.2e-16 ***
## SSF     1     16.077 143.96 -60.423  24.7654 1.408e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#All F values are at least 3 at this point, so our final model includes Ht, Wt, LBM and SSF

winner.url = "http://users.stat.ufl.edu/~winner/data/clinton1.dat"
tmp.url = paste0(winner.url, "clinton1.data")
st = c(1, 39, 46, 53, 62, 70, 78, 86, 95, 106, 115)
en = c(38, 45, 52, 61, 69, 77, 85, 94, 105, 114, 124)
wds = diff(c(st, en[length(en)]))
clinton = read.fwf(winner.url, widths = wds, strip.white = T)
colnames(clinton) = c("country", "vote", "age", "savings", "income", "poverty", "veterans", "female", "other")
attach(clinton, warn.conflicts = F)
plot(clinton)

```



### Question 3

```

attach(clinton, warn.conflicts = F)
fitmod <- lm(vote ~ 1, data=clinton)
indep.vars <- ~ age + savings + income + poverty + veterans + female + density + nursing + crime
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ 1
##          Df Sum of Sq    RSS   AIC F value    Pr(>F)
## <none>            280213 12551
## age      1     947 279266 12544  9.1637  0.002492 ***
## savings  1    5508 274705 12499 54.1757 2.421e-13 ***
## income   1   14139 266074 12413 143.5847 < 2.2e-16 ***
## poverty  1   70344 209869 11771 905.6622 < 2.2e-16 ***
## veterans 1   3171  277042 12522 30.9290 2.939e-08 ***
## female   1   20246  259967 12350 210.4311 < 2.2e-16 ***
## density  1    9539  270674 12459  95.2223 < 2.2e-16 ***
## nursing  1   2454  277759 12529  23.8702 1.091e-06 ***
## crime    1   4078  276135 12513  39.9022 3.112e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + poverty)
add1(fitmod, indep.vars, test = "F")

## Single term additions

```

```

## Model:
## vote ~ poverty
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      209869 11771
## age       1     138.9 209730 11771   1.7887  0.181203
## savings   1     374.8 209494 11768   4.8328  0.028008 *
## income    1     3236.8 206632 11731  42.3096 9.250e-11 ***
## veterans  1     548.7 209320 11766   7.0799  0.007842 **
## female    1     8478.0 201391 11662 113.7045 < 2.2e-16 ***
## density   1     9749.0 200120 11644 131.5810 < 2.2e-16 ***
## nursing   1     1225.1 208644 11757  15.8600 7.001e-05 ***
## crime     1     2263.2 207606 11744  29.4446 6.264e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + density)
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ poverty + density
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      200120 11644
## age       1     355.7 199764 11642   4.8071  0.028428 *
## savings   1     1245.5 198874 11630  16.9091 4.038e-05 ***
## income    1     460.9 199659 11640   6.2327  0.012600 *
## veterans  1     526.5 199593 11639   7.1220  0.007660 **
## female    1     6533.2 193586 11557  91.1207 < 2.2e-16 ***
## nursing   1     549.0 199571 11639   7.4279  0.006463 **
## crime     1     66.4 200053 11646   0.8955  0.344065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + female)
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ poverty + density + female
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      193586 11557
## age       1     1.18 193585 11559   0.0164  0.89815
## savings   1    2829.06 190757 11519  40.0280 2.921e-10 ***
## income    1     23.88 193563 11558   0.3330  0.56393
## veterans  1     254.97 193332 11555   3.5595  0.05931 .
## nursing   1    1262.50 192324 11541  17.7174 2.646e-05 ***
## crime     1     31.02 193555 11558   0.4326  0.51079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + savings)
add1(fitmod, indep.vars, test = "F")

```

```

## Single term additions
##
## Model:
## vote ~ poverty + density + female + savings
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      190757 11519
## age       1     333.65 190424 11516  4.7273 0.029774 *
## income    1     393.39 190364 11515  5.5755 0.018284 *
## veterans  1    1051.19 189706 11506 14.9500 0.000113 ***
## nursing    1     198.88 190559 11518  2.8157 0.093459 .
## crime      1      8.76 190749 11521  0.1239 0.724833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + veterans)
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ poverty + density + female + savings + veterans
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      189706 11506
## age       1     14.92 189691 11508  0.2122 0.645110
## income    1     443.52 189263 11502  6.3202 0.011995 *
## nursing   1     520.83 189185 11501  7.4249 0.006474 **
## crime     1      3.13 189703 11508  0.0446 0.832830
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + nursing)
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ poverty + density + female + savings + veterans + nursing
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      189185 11501
## age       1    132.732 189053 11501  1.8928 0.16900
## income    1    272.830 188913 11499  3.8936 0.04857 *
## crime     1    77.805 189108 11502  1.1092 0.29235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . + income)
add1(fitmod, indep.vars, test = "F")

## Single term additions
##
## Model:
## vote ~ poverty + density + female + savings + veterans + nursing +
##       income
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>      188913 11499
## age       1    150.28 188762 11498  2.1455 0.14310

```

```

## crime    1    204.09 188708 11498  2.9146 0.08789 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##final model includes poverty, density, female, savings, veterans, nursing and income.

```

## Question 4

```

fitmod <- lm(vote ~ age + savings + income + poverty + veterans + female + density + nursing + crime, da)
drop1(fitmod, test="F")

## Single term deletions
##
## Model:
## vote ~ age + savings + income + poverty + veterans + female +
##      density + nursing + crime
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>           188611 11498
## age     1       97 188708 11498   1.3894 0.2386186
## savings 1      2732 191343 11535  39.0199 4.86e-10 ***
## income   1       395 189006 11502   5.6411 0.0176138 *
## poverty  1      34250 222861 11948 489.2076 < 2.2e-16 ***
## veterans 1       885 189496 11509  12.6342 0.0003853 ***
## female   1      6981 195592 11595  99.7076 < 2.2e-16 ***
## density   1      6562 195173 11589  93.7258 < 2.2e-16 ***
## nursing   1      546 189158 11504   7.8052 0.0052464 **
## crime     1      151 188762 11498   2.1580 0.1419496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . - age)
drop1(fitmod, test="F")

## Single term deletions
##
## Model:
## vote ~ savings + income + poverty + veterans + female + density +
##      nursing + crime
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>           188708 11498
## savings  1      2677 191386 11534  38.2356 7.221e-10 ***
## income    1       399 189108 11502   5.6998 0.017035 *
## poverty   1      34181 222890 11946 488.1504 < 2.2e-16 ***
## veterans  1      1386 190095 11516  19.7962 8.965e-06 ***
## female    1      7497 196206 11601 107.0715 < 2.2e-16 ***
## density   1      6553 195262 11588  93.5878 < 2.2e-16 ***
## nursing   1      466 189174 11502   6.6505 0.009965 **
## crime     1      204 188913 11499   2.9146 0.087894 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fitmod <- update(fitmod, . ~ . - crime)
drop1(fitmod, test="F")

## Single term deletions

```

```

## 
## Model:
## vote ~ savings + income + poverty + veterans + female + density +
##      nursing
##          Df Sum of Sq   RSS   AIC F value    Pr(>F)
## <none>           188913 11499
## savings     1     2534 191447 11533  36.1629  2.06e-09 ***
## income      1      273 189185 11501   3.8936   0.04857 *
## poverty     1    34688 223601 11952 495.0397 < 2.2e-16 ***
## veterans    1    1351 190264 11516  19.2840  1.17e-05 ***
## female      1    7418 196330 11601 105.8613 < 2.2e-16 ***
## density     1    6461 195374 11588  92.2072 < 2.2e-16 ***
## nursing     1      350 189263 11502   4.9969   0.02547 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

*#All F values are at least 3 at this point, so our final model includes savings, income, poverty, veterans, and nursing.*

## Question 5

Unequal sample sizes can lead to: Unequal variances between samples, which affects the assumption of equal variances in tests like ANOVA, violating assumptions we could make. There are definitely unequal variance throughout the different values of independent variables. To be specific, in an extreme cases, a mean of sample size of 2 will have larger variance than a mean of sample size of 10000. Relating to this, in order to satisfy Gauss-Markov, we would need to control for equal variance, and thus should add population as an estimator for our models. This data would likely violate the constant variance assumption.